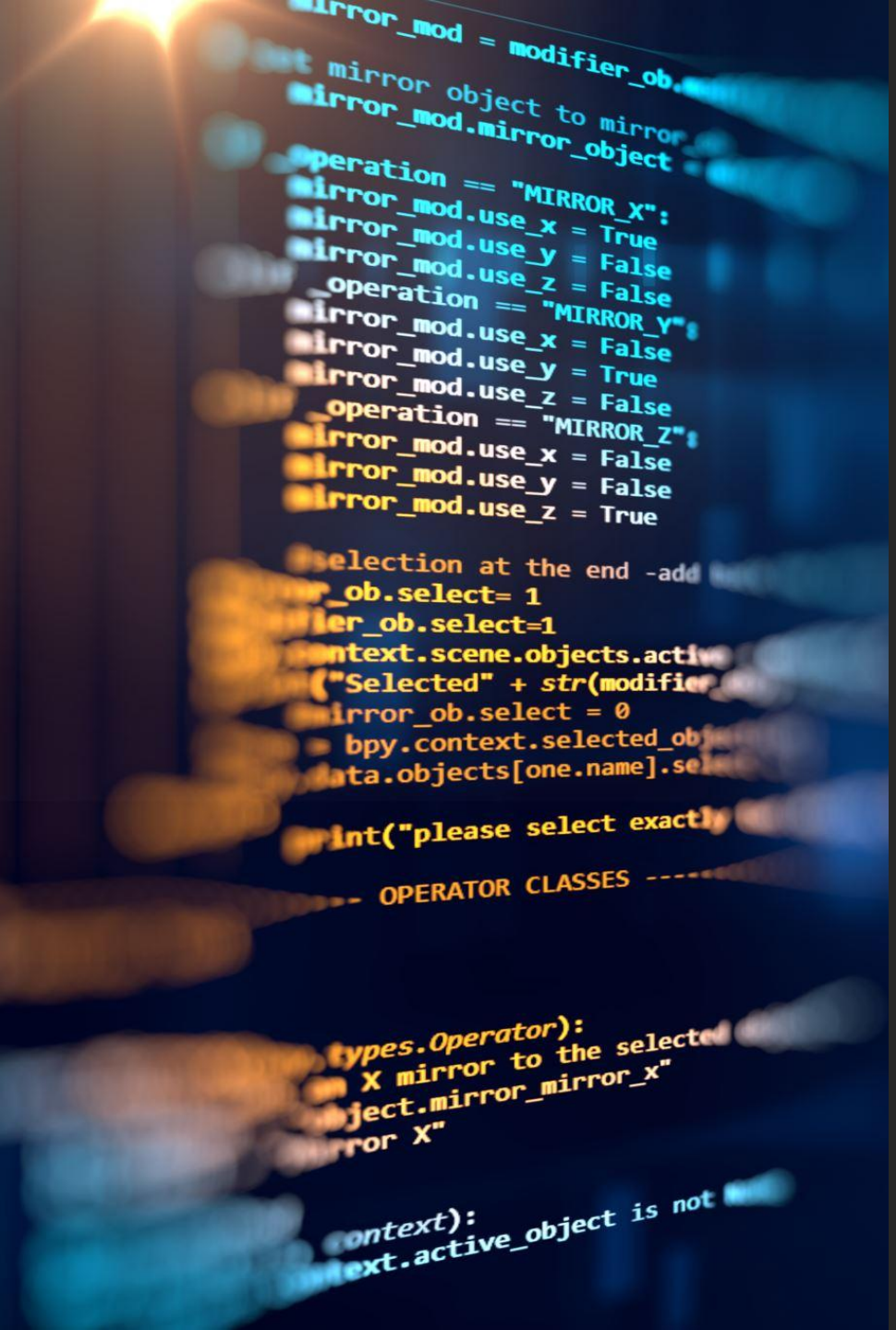
Natural Language Processing (NLP)

Analisis Sentimen Kebijakan Pemerintah

By: Wahyu Calvin



Outline:

- Pemanfaatan Big Data di BPS
- NLP dan Analisis Sentimen
- Praktik (*hands-on session*)

PEMANFAATAN BIG DATA DI BPS

BADAN PUSAT STATISTIK (BPS)

Badan Pusat Statistik (BPS) adalah Lembaga Pemerintah Non Kementerian yang bertanggung jawab langsung kepada Presiden.

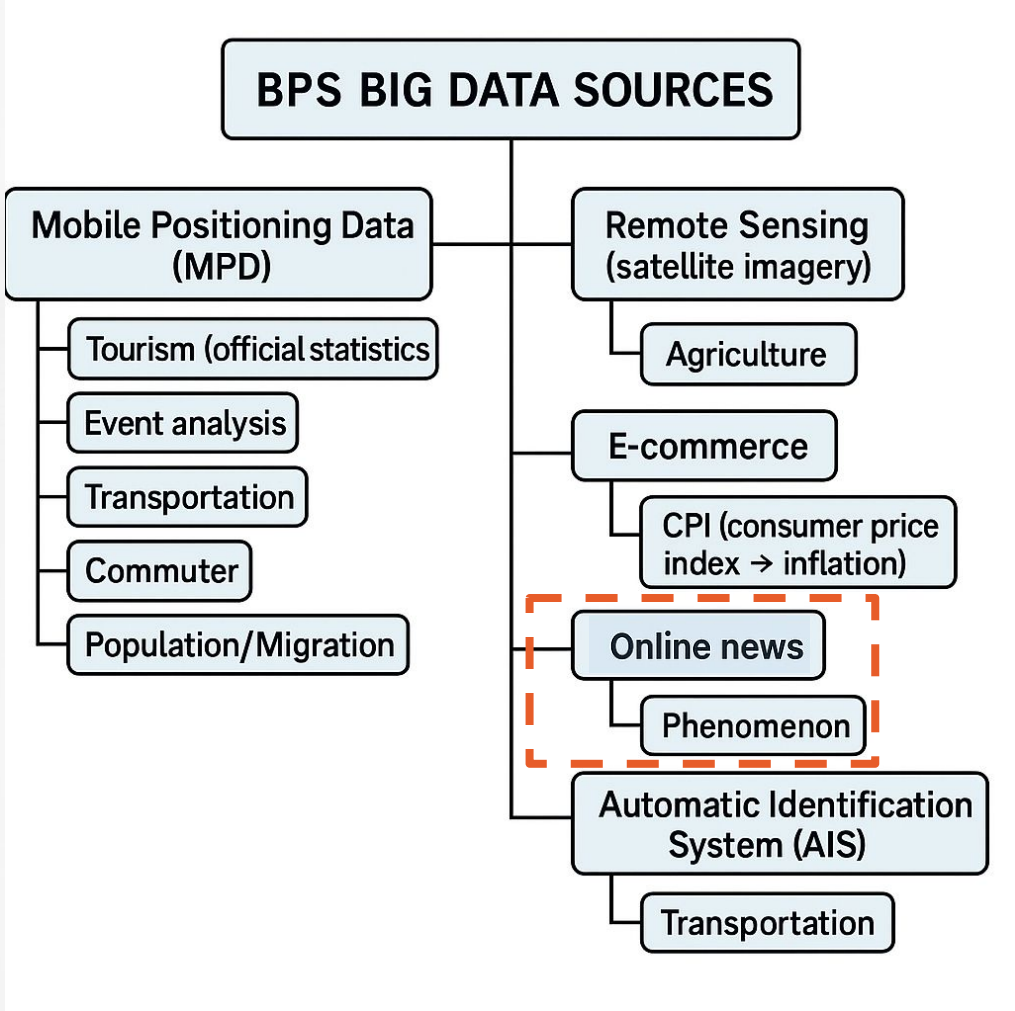
BPS bertugas dalam melaksanakan pemerintahan di bidang statistik sesuai peraturan perundang-undangan. Hasil statistik yang diselenggarakan oleh BPS diumumkan dalam Berita Resmi Statistik (BRS) secara teratur dan transparan agar masyarakat dengan mudah mengetahui dan mendapatkan data yang diperlukan.

BEBERAPA DATA YANG DIHASILKAN BPS

Sumber Data BPS:

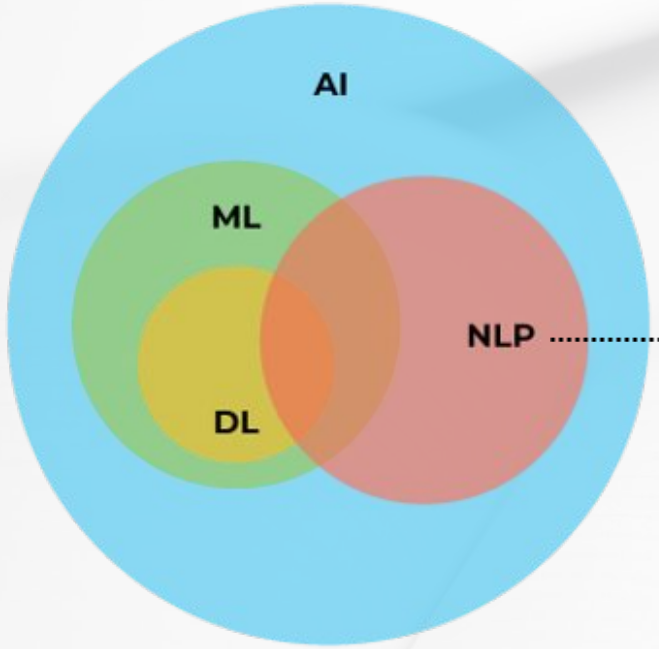
- Sensus
- Survey
- Registrasi
- *Big Data** (sumber data baru)

PEMANFAATAN BIG DATA DI BPS



NATURAL LANGUAGE PROCESSING (NLP) & SENTIMENT ANALYSIS

NATURAL LANGUAGE PROCESSING (NLP)



- Artificial Intelligence
- Machine Learning
- Language Processing
- Deep Learning

Natural Language Processing (NLP) adalah bidang yang mempelajari metode, algoritma, dan representasi yang memproses input ataupun output yang berupa *bahasa alami*, baik berupa teks maupun suara.

- Text Summarization
- Information Extraction
- Information Retrieval
- Sentence Similarity
- **Text Categorization** → **Sentiment Analysis**
- Sequence to Sequence (generation)

ANALISIS SENTIMEN

Analisis Sentimen (*sentiment analysis*) merupakan analisis untuk menentukan nada emosional dari penulisnya, atau juga dapat didefinisikan sebagai proses memahami dan mengelompokkan emosi (biasanya dikelompokkan sebagai positif, negatif, atau netral).

Contoh:

- “Layanan masyarakat sangat lambat dan sop tidak jelas” ⇒ **Negatif**
- “Pemerintah membuka pendaftaran CPNS mulai minggu depan” ⇒ **Netral**
- “Saya puas dengan pelayanan cepat saat membuat paspor” ⇒ **Positif**

ANALISIS SENTIMEN

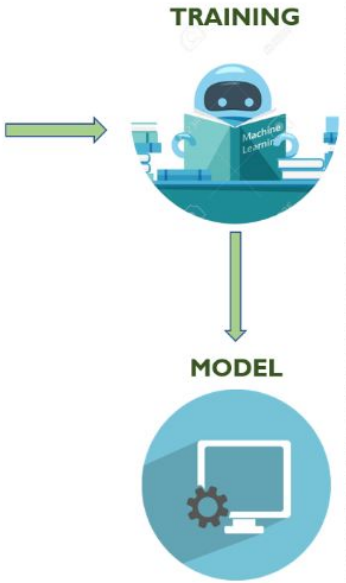
LEXICON BASED

KAMU KEREN +4	POSITIF
KAMU TIDAK KEREN -4	NEGATIF
KAMU SANGAT KEREN +6 +4	POSITIF
KAMU KEREN TAPI JELEK -4	NETRAL
KAMU SANGAT JELEK -6	NEGATIF

- **SENTIMEN POSITIF**
  Skor positif terbesar > Skor negatif terbesar
- **SENTIMEN NEGATIF**
  Skor negative terbesar > Skor positif terbesar
- **SENTIMEN NETRAL**
  Skor positif terbesar = Skor negatif terbesar

MACHINE LEARNING BASED

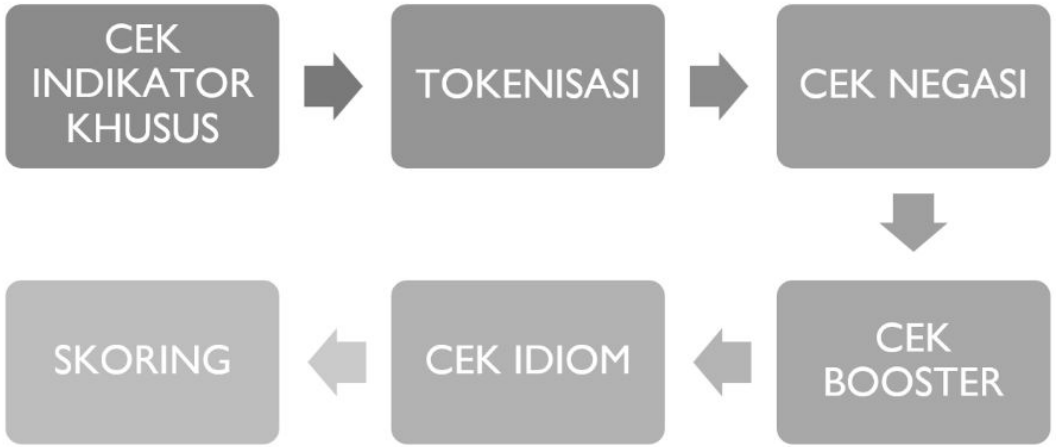
Sentiment	Text Tweet
negative	Jelek filmnya... apalagi si ernest gak mutu bgt actingnya... film sampah
negative	Film king Arthur ini film paling jelek dari seluruh cerita King Arthur
negative	@beexkuanlin Sepanjang film gwa berkata kasar terus pada bapaknya
negative	Ane ga suka fast and furious..menurutku kok jelek ya tu film
positive	10. Kalaupun mau bikin film horror, film Danur bisa jadi rujukan .. ringan tapi menakutkan
positive	Banyak moral yg bisa dipetik dari film wonder woman. Salah satunya adalah menjaga auratmu dr cowok hidung belang ?
positive	Berikut 5 film horor Indonesia berpendapatan tinggi dalam 10 tahun terakhir. Dapatkan tiket @DanurMovie di http://bit.ly/belidanur
positive	Trust me. #GetOut itu tipe film paling nyaman dinikmati saat kita gak tau sama sekali filmnya mau kemana. Kalau sudah baca sana-sini ya beda



ANALISIS SENTIMEN

LEXICON BASED

PENDEKATAN LEKSIKON SENTISTRENGTH

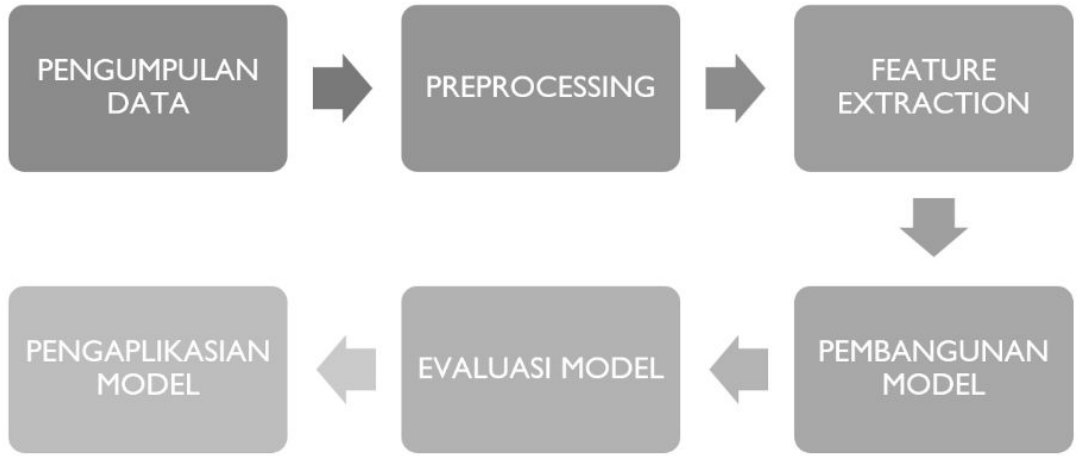


Algoritma : SentiStrength Indonesia

Penelitian Terkait: https://doi.org/10.22146/ijccs.16625

MACHINE LEARNING BASED

PEMBANGUNAN MODEL ANALISIS SENTIMEN



Algoritma : Neural Network (*Deep Learning*)

Penelitian Terkait: https://doi.org/10.13106/jidb.2022.vol13.no6.9

ANALISIS SENTIMEN

TRANSFER LEARNING

Transfer Learning teknik pembelajaran mesin (*machine learning*) yang memanfaatkan pengetahuan yang telah dipelajari oleh model pada satu tugas untuk meningkatkan kinerja model pada tugas yang terkait atau berbeda. Dengan kata lain, transfer learning adalah penggunaan kembali model yang sudah dilatih (*pre-trained* model) sebagai titik awal untuk melatih model baru pada tugas yang berbeda.

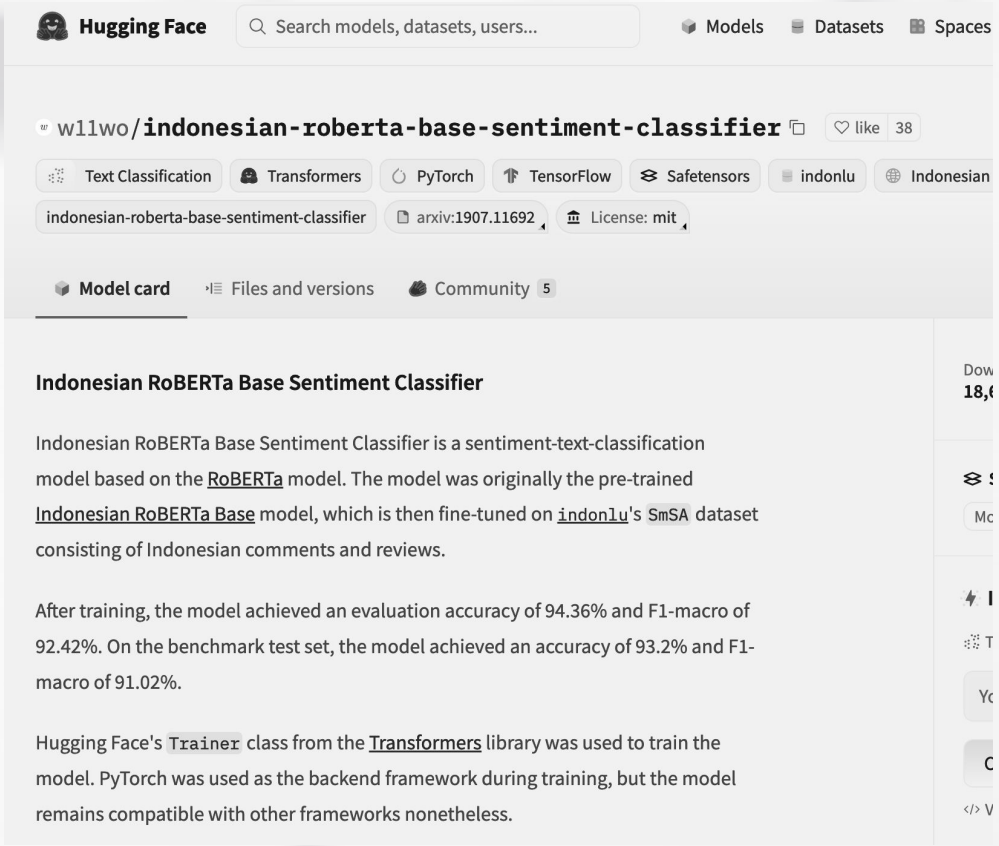
Generasi	Teknik	Kekurangan
1	Rule-Based	Terbatas kosakata
2	Machine Learning (SVM, NB)	Perlu feature engineering
3	Deep Learning (LSTM, CNN)	Perlu data besar
☑ 4	**LLM & Transfer Learning** (RoBERTa, BERT)	🔥 Akurasi tinggi, siap pakai

ANALISIS SENTIMEN

TRANSFER LEARNING

Model: **w11wo/indonesian-roberta-base-sentiment-classifier**

- Pre-trained dari RoBERTa base
- Dikembangkan oleh komunitas NLP Indonesia
- Di-host di: HuggingFace Model Hub
- Label: positif, negatif, netral



HANDS-ON SESSION

HANDS-ON SESSION

https://github.com/calvinrev/iprakom-scraping-sentiment.git


THANK YOU